Claims

Having thus described our invention, what we claim as new and desire to secure by Letters Patents is.

Sub A1

1. In a memory structure having a plurality of memory levels, a forward and return path for interconnecting said memory levels and a processor, a method of transferring requests from said processor on said forward paths and responses to said requests on said return path to said processor, said method comprising:

transmitting each request from said processor to each of said levels;

if a request from a memory level of said levels is a read to the memory of latter said memory level, then transmitting a response to latter said read request on said return path along with another response from internal buffers of latter said memory level;

if a request from a memory level of said memory levels is a write request to the memory of latter said memory level, then transmitting two responses from internal buffers of latter said memory level on said return path;

if a request from a memory level of said memory levels is targeted to a higher memory level of said memory level, then transmitting two responses from internal buffers of latter said memory level on said return path; and

if a request from a memory level of said memory levels is targeted to a lower memory level of said memory level, then transmitting one response from internal buffers of latter said memory level on said return path.

2. A method as recited in claim 1, wherein at most three responses are stored in said internal buffer of any of said memory levels.

3. A memory structure for receiving requests to and transmitting responses from a memory in said structure, respectively, said structure comprising:

an ordered set of memory levels, each memory level having a controller and memory unit, which unit is a portion of said memory;

a forward path for transmitting said requests to said memory levels, starting from lower of said levels and proceeding to a higher of said levels;

and a return path for transmitting responses to said requests from higher of said levels to lower of said levels, wherein each controller at each memory level transmits responses to said requests on said return path from each level to a buffer in said controller at a lower level in accordance with the following algorithm:

a. if a request at any one of said levels is a read to said one level, then transmitting a first response to latter said request on said return path, along with a second response from internal buffers of said controller of said one level, to buffers in a controller of a memory level which is lower than said one level,
b. if a request at any one of said levels is a write to said one level, then transmitting two responses from internal buffers of said controller of said one level on said return path to buffers in a controller of a memory level which is lower than said one level,
c. if a request at any one of said levels is targeted to a higher of said levels, then transmitting two responses from internal buffers of said controller of said one level on said return path to buffers in a controller of a memory level which is lower than said one level, and
d. if a request at any one of said levels is targeted to a lower of said levels, then transmitting one response to a request from internal buffers of said controller of said one level on said return path to buffers in a controller of a memory level which is lower than said one level.

4. A memory structure as recited in claim 3, wherein each of said controllers on said return path has buffers for storing at most three of said responses.

5. A memory structure as recited in claim 3, wherein said return path has twice the bandwidth of said forward path.

6. A memory structure as recited in claim 3, wherein each of said memory units has internal logic for copying data between one of said memory units and a request as specified therein, and forwarding said one modified request to controllers on said return path.